Detection and Analysis of DNS Dependence

Fang Qiang[1], Gong Jian[1 2],Yang Wang[1 2]

1.School of Computer Science & Engineering, Southeast University, Nanjing, 211189;

2. Jiangsu Provincial Key Laboratory of Computer Network Technology, Southeast University, Nanjing, 211189

1.qfang@njnet.edu.cn,2.jgong@njnet.edu.cn,3.wyang@njnet.edu.cn

Abstract: Detect DNS dependence is very important for a network in DNS security. This paper takes the DNS packets passing through border router as the data sources, and take the number of the users' IPs accessing a DNS as the dependence degree metric, then clustering the dependence metric with k-mean clustering algorithm to find the DNS with high dependence.

Key words: DNS Dependence Border Router k-mean Clustering

DNS 依赖性的检测与分析

方强[1]，龚俭[1 2]，杨望[1 2]

1．计算机科学与工程学院 东南大学，南京，211189，

2．江苏省计算机网络技术重点实验室 东南大学，南京，211189

1.qfang@njnet.edu.cn,2.jgong@njnet.edu.cn,3.wyang@njnet.edu.cn

摘要: 检测一个网络的 DNS 依赖性对于保护该网络使用 DNS 的安全有着重要的意义。本文提出采集经过网络边界路由器的 DNS 报文获取完整数据来源，使用 DNS 被访问的 IP 数作为依赖性指标，并且使用 k-mean 聚类算法按照依赖性指标进行聚类，找出依赖性强的 DNS。

关键字: DNS 依赖性 网络边界路由器 k-mean 聚类

1.引言

DNS[4](域名系统，Domain Name System)是因特网的一项核心服务，它作为可以将域名和 IP 地址相互映射的一个分布式数据库，能够使人更方便的访问互联网，而不用去记住能够被机器直接读取的 IP 数串。，由于 DNS 的存在使得互联网的广泛使用成为可能。由于 DNS 在互联网中的重要作用，使得 DNS 的安全对于互联网的安全有着重要的意义。一个 DNS 出现故障将会导致大量的域名无法解析，从而导致大规模的网络故障。

当前 DNS 安全的研究主要关注是防范 DNS 欺骗以及防范对于 DNS 的攻击。，目前还很少有文献从网络管理员的角度去说明如何去保证一个网络的 DNS 安全;一个网络的 DNS 安全既包括该网络内部用户所使用的外部 DNS 的安全，也包括外部用户使用网络内部 DNS 的安全。达成上述目的的首要工作是了解网络外部用户对于内部 DNS 的依赖性以及网络内部用户对于外部 DNS 的依赖性。本文介绍了一种检测 DNS 依赖性的方法，该方法对经过网络边界路由器的 DNS 报文进行解析和统计，并对统计的结果进行聚类分析，从而实现对于 DNS 依赖性的检测。

全文安排如下:第 2 节介绍了依赖性检测的模型；第 3 节介绍了实验的过程以及获得的数据；第四节对本文进行了总结。

2．依赖性检测模型

2.1．基本思想

一个网络的 DNS 依赖性可分为外部用户对于内部 DNS 的依赖性和网络内部用户对于外部 DNS 的依赖性。

想要检测 DNS 的依赖性，首先就需要知道内部用户使用的外部 DNS 和外部用户使用的 DNS 的域名(或者是 IP)，网络内部用户访问外部 DNS 以及外部用户访问内部 DNS 时所发送的 DNS 报文一定会经过网络边界路由器。因而抓取通过网络边界路由器的 DNS 报文并对其进行解析将能够获取很多重要信息，包括该报文的源宿地址，请求解析的域名，解析该域名的 DNS 的域名及地址，可以从中选取所需的数据用于分析。

在获取到检测所需的数据之后，需要找到一个指标用来评判用户对于各个 DNS 的依赖性(即对各个 DNS 的依靠的程度)。有两个候选指标可以用来使用：一个 DNS 被访问的次数和一个 DNS 被访问的 IP 数。记录一个 DNS 被访问次数往往不能准确，因为大多数用户都会设置 DNS 缓存，用户在访问一次 DNS 获取到所需的 IP 地址之后将会在一点时间内使用 DNS 缓存而不是重新访问 DNS，因而会影响到数据的准确

性。采用 DNS 被访问的 IP 数作为指标则不会出现该问题，无论一个 IP 访问过该 DNS 的次数是多少，该 IP 都会被记录下来。

在获取到所需数据以及评判标准之后，就需要从大量的 DNS 中找出对于该网络来说最重要的那些 DNS，需要确定的数据包括重要的 DNS 的数量以及它们的域名(或者是 IP)，本文采用了 k-mean 聚类算法，按照 DNS 被访问的 IP 数量进行分类，选择其中数量最多的一类或者是几类作为重要的 DNS。

2.2．数据的获取及统计

要获取用户使用的 DNS 服务器以及获取访问每一个 DNS 服务器的 IP 地址，需要解析通过网络边界路由器的 DNS 报文，DNS 报文的格式如下图所示(包括 IP 头以及 UDP 报头):

以太报文头
IP 报文头
UDP 报文头
DNS 报文头
DNS 报文查询段
DNS 报文应答段
DNS 报文授权段
DNS 报文附加段

Figure 1.complete DNS packet format

图 1. DNS 完整报文格式

在 DNS 报文头中能够获取该 DNS 报文的类型，筛选出其中的 DNS 回复报文作为待分析报文，从 DNS 报文的授权段可以获取解析被请求解析域名的 DNS 服务器的域名(注：DNS 服务器也有自己的域名),同时从 IP 报文头中获取报文的宿地址，即用户地址。通过收集大量报文可以获取大量 DNS 服务器域名以及访问这些 DNS 的用户 IP。

2.3．DNS 服务器的依赖性分类

进行依赖性分类的目的在于找出用户依赖性高的那些 DNS，对 DNS 按照被访问的 IP 个数进行分类，找出的被访问 IP 个数多的那些 DNS 就是用户依赖性高的 DNS。

在分类中采用了 k-mean[2]聚类算法。具体实现该算法的方式如下：

1. 从全部的数据中任意选择 k 个数据作为初始聚类中心；
2. 对于所剩下其它数据，则根据它们与这些聚类中心的距离(即所对应的 IP 地址个数之间的差值)，分别将它们分配给与其距离最小的聚类；然后再计算每个所获新聚类的聚类中心（该聚类中所有对象的均值）；
3. 不断重复这一过程直到各个聚类所包含的数据对象不再改变为止。

K-mean 聚类具有以下特点：各聚类本身尽可能的紧凑，而各聚类之间尽可能的分开。

4. 实验以及结果

本文使用 CERNET 江苏省网作为实验对象。实验环境为：Linux version 2.6.9-42.ELsmp (Red Hat 3.4.6-2))，内存 4G，双核 CPU、 主频: 2400.587MHz；通过嗅探该主机的 eth1 接口可以获取通过江苏省网边界的 DNS 报文。

为实现上述算法，获取所需数据，本文利用 C 语言[3]设计了 dns_cpature，statistics_ips，以及 k_mean 三个程序，各个程序的功能如下：

1) dns_capture 程序被开发用来从江苏省网边界抓取 DNS 报文，并且从中获取所需的信息：DNS 服务器的域名以及使用该 DNS 的用户 IP，对每一分钟内获取的信息整理之后存储到文件中，文件中数据存储的格式为：

DNS 域名 IP 数
IP_1
。。。
IP_n

Figure 2.data store format

图 2.采集数据保存格式

dns_capture 程序使用了 Libpcap[4]库抓取报文，采用 libbind[5]库解析 DNS 报文，dns_capture 在实验主机上运行 10 天，抓取了从 2010.8.2-2010.8.12 经过江苏省网边界的 DNS 报文。

2) statistics_ips 程序用于统计 dns_capture 程序获取的 DNS 报文信息，获取每一个 DNS 被访问的 IP 个数。并将所有的 DNS 按照 IP 地址数目从大到小排序，在实验中一共获取了 24654 个江苏省网内部的 DNS 服务器，316983 个江苏省网外部的 DNS 服务器。

3) k_mean 程序对 statistics_ips 所获得的 DNS 按照被访问的 IP 的数量进行聚类。通过设置将分类的

个数设为 4、6、7,可以看到如下的结果：

a) 若将 DNS 服务器分为 4 大类(即 K=4)，结果如下：

Table 1. the clustering situation of outer DNS when the data is divided into 4 classes

表 1.将数据分为 4 类时外部 DNS 的聚类情况

聚类中心(DNS 服务器关联 IP 数)	聚类大小(DNS 服务器个数)
858	316668
2574.5	194
8296.5	105
32150.	14

Table 2. the clustering situation of inner DNS when the data is divided into 4 classes

表 2.将数据分为 4 类时内部 DNS 的聚类情况

聚类中心(DNS 服务器关联 IP 数)	聚类大小(DNS 服务器个数)
1129.50	24506
3566.50	77
34072	11
10865.50	58

b) 若将 DNS 服务器分为 6 大类(即 K=6)，结果如下：

Table 3. the clustering situation of outer DNS when the data is divided into 6 classes

表 3. 将数据分为 6 类时外部 DNS 的聚类情况

聚类中心	聚类大小(DNS 服务器个数)
95	315661
285.5	606
858	401
2574.5	194
8296.5	105
32150	12

Table 4. the clustering situation of inner DNS when the data is divided into 6 classes

表 4. 将数据分为 6 类时内部 DNS 的聚类情况

聚类中心	聚类大小(DNS 服务器个数)
119.50	24316
363.50	115
1129.5	75
3566.5	77
34073	9
10865.5	58

c) 若将 DNS 服务器分为 7 大类(即 K=7)，结果如下：

Table 5. the clustering situation of outer DNS when the data is divided into 7 classes

表 5. 将数据分为 7 类时外部 DNS 的聚类情况

聚类中心(访问一个 DNS 的 IP 数)	聚类大小
31.50	313116
95	2545
285.50	606
858	401
2574.5	194
8296.50	105
32150	11

Table 6. the clustering situation of inner DNS when the data is divided into 7 classes

表 6. 将数据分为 7 类时内部 DNS 的聚类情况

聚类中心(访问一个 DNS 的 IP 数)	聚类大小
39.00	24215
119.50	101
363.50	115
1129.50	75
3566.50	77
34072	8
10865.5	58

从表 1-6 中可以清楚的看到，江苏省网内部的 DNS 的总数为 24654，聚类中心最大的两个类别与其他的类别之间的聚类中心距离较远，它们可以作为江苏省网外部用户依赖性高的内部 DNS。同时，它们的数量基本稳定在 70 左右，占内部 DNS 总数的 2.83‰。同样可以看到江苏省网外部的 DNS 的总数为 316983，其中聚类中心最大的两个类别可以作为江苏省网内部用户依赖性高的外部 DNS。同时，它们的数量基本稳定在 130 左右，大概占外部 DNS 总数的 0.41‰。因此，使用 k-mean 聚类算法可以将 DNS 按照依赖性进行有效的分类。

总结

DNS 是互联网中不可或缺的重要设备，对于一个

网络来说保护其网络内部用户使用的网络外部 DNS 以及网络外部用户使用的内部 DNS 的安全是非常重要的工作，检测网络的 DNS 依赖性是实现上述目的的重要步骤，找出对于一个网络来说依赖性强的 DNS 将有助于对 DNS 安全的保护。本文提出 DNS 依赖性分为网络外部用户对于内部 DNS 的依赖性和网络内部用户对于外部 DNS 的依赖性。通过抓取经过网络边界路由器的 DNS 报文能够有效了解内部用户对于外部 DNS 以及外部用户对于内部 DNS 的访问情况，使用一个 DNS 被访问的 IP 数作为依赖性指标，并且使用 k-mean 聚类算法按照 DNS 被访问的 IP 数作为聚类数据，可以对 DNS 按照依赖性进行有效的分类，从而找出依赖性高的 DNS。

References (参考文献)

[1] W.Recard Stevens ，TCP/IP 详解，卷一：协议，中文版，机械工业出版社

[2] J. Han and M. Kamber, Morgan Kaufmann ,Data Mining: Concepts and Techniques

[3] 于明俭、陈向阳、方汉等，Linux 程序设计权威指南，机械工业出版社

[4] Paul Albitz、Cricket Liu，DNS and BIND, 4th Edition，O'Reilly

[5] 阿美，Pcap 程序设计，URL：http://broker.dhs.org/pcap.htm

方强
东南大学计算机科学与工程学院
江苏南京，211189
电话：13401961483
Email: qfang@njnet.edu.cn
简历：
东南大学计算机科学与工程学院硕士研究生，研究领域：网络安全

龚俭
东南大学计算机科学与工程学院 江苏省网络术重点实验室
江苏南京，211189
Email: jgong@njnet.edu.cn
简历：
东南大学计算机科学与工程系教授，博士生导师； CERNET 华东(北)地区网络中心主任，主要研究领域：网络行为学，网络安全

杨望
东南大学计算机科学与工程学院 江苏省网络术重点实验室
江苏南京，211189
Email: wyang@njnet.edu.cn
简历：
东南大学计算机科学与工程系讲师，博士；主要研究领域：网络管理与网络安全